THE USE OF THE TWO DIMENSIONAL CONTINUOUS WAVELET TRANSFORM FOR CLASSIFICATION OF TARGETS IN FLIR IMAGERY

A THESIS

SUBMITTED TO THE FACULTY OF CLARK ATLANTA UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

BY

DENIS SEMWOGERERE

DEPARTMENT OF PHYSICS

ATLANTA, GEORGIA

MAY 1998

ABSTRACT

SEMWOGERERE, DENIS B. B.S., UNIVERSITY OF LESOTHO, 1992

THE USE OF THE TWO DIMENSIONAL CONTINUOUS WAVELET TRANSFORM FOR CLASSIFICATION OF TARGETS IN FLIR IMAGERY

Advisor: Dr. Romain Murenzi

Thesis Dated May, 1998

Two dimensional continuous wavelet transforms have been used to study a wide variety of physical problems especially in the analysis of images. They are used in this particular application for the detection of discontinuities, which include edges, filaments, contours and boundaries between areas of different luminosity.

In this study we investigate the use of a continuous wavelet transform partial energy density as a features detector in developing an algorithm for classification of targets in forward looking infrared imagery. Our algorithm specifically uses features that result from computing the partial energy densities of targets at all relevant scales and orientations and classification of those targets using a distance based classifier.

We have found in our analysis that the partial energy density can extract target features that result in an overall improvement in classification in our algorithm.

Table of Contents

Acknowledgements

I would like to extend my gratitude to the Center for Theoretical Studies of Physical Systems for supporting me financially at CAU, and helping me pursue further education elsewhere. Special thanks to Dr. Romain Murenzi, my advisor, and Dr. Lance Kaplan for their enormous contributions to this masters research project. Thank you to my committee members, who include Dr. Lonzy Lewis Chair, Department of Physics and Dr John Hurley for their constructive comments and special assistance. I have asked them to go "above and beyond" on numerous occasions. Many thanks to Ms. Davida Johnson and Dr. Kamesh Namuduri. Thanks to the physics department for affording me this opportunity. Special recognition to my father, Dr. Fred Semwogerere without whom none of this would have been possible. All the sacrifices and hardships that he has endured and continues to weather for the good of the family are greatly appreciated. And last, but certainly not least, I thank God for the good fortune He has bestowed upon me.

List of Figures

List of Tables

List of Abbreviations

ARP	Angular Resolving Power
ATR	Automatic Target Detection and Recognition
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
FFT	Fast Fourier Transform
FLIR	Forward Looking Infrared
MCWT	Multidimensional Continuous Wavelet Transform
PDF	Probability Density Function
SAR	Synthetic Aperture Radar
SRP	Scale Resolving Power

Chapter 1
Introduction

This thesis is part of a Department of Defense funded automatic target detection and recognition (ATR) project. Our aim is to develop an efficient and accurate ATR algorithm using the continuous wavelet transform (CWT) to build a features detector for eventual input into a convolutional neural network (CNN) [1]. To understand the motivation behind the use of the CWT it is instructive to have some knowledge of ATR.

ATR is the use of computer processing power to detect and recognize target signatures in sensor data. Common sensors include forward-looking infrared (FLIR) imagery, synthetic-aperture radar (SAR), multispectral optical millimeter wave radar, and laser radar. In this study, however, we restrict our work to FLIR images.

ATR allows analysis of data at rates that make it possible to process the huge amounts of data that result from the deployment of new systems. These systems are expected to become operational within the next few years.

The fundamental problem of ATR is to detect and recognize objects of interest (targets) obscured by clutter from sensors that add noise to the signal. Clutter refers to objects such as buildings, cars, trucks, grass, trees, rocks that are not the primary targets. Typically, images are dominated by clutter because targets are generally sparse compared to their environment. Noise includes electronic noise in the sensor as well as inaccuracies introduced in the computations by a signal processor.

The development of a successful ATR system is made difficult by the large amount of target variability in the sensor data. In FLIR images, for example, a tank can be hotter or colder than its background causing positive or negative thermal contrast. Variations can result from translation, rotation, and scaling of targets, as well as sensors which can introduce noise into the image. Additionally, there is the problem of different targets having similar distinguishing characteristics (signatures), i.e., a particular target can have different signatures. For example the turret of a tank can be rotated to any position and its gun barrel can take on a range of elevation angles which make it appear different from identical tanks at the same range, rotation and elevation angle. Further complications are introduced by variations in the background clutter which can be anything from a featureless desert, to a tree line, to a highly cluttered urban area. Adding to the difficulty is the limited training and testing data that would allow for the capture of these variations.

Given the complexity of ATR, it is important to have the ability to optimally extract the essential features of a target. In this regard, efficient data representations are required in which the important features are both compactly and clearly represented which can enhance ATR. Since both detection and identification are important, multidimensional data representations and analysis techniques, such as the CWT, are highly desirable. Murenzi et al [1,2] suggested a three-step algorithm for ATR and proposed that the efficiency and robustness may be the result of the 2D CWT mapping of images in the space of the characteristic features of the targets: position, scale, orientation and anisotropy. Scale dependence allows for sensitivity in sensor resolution and determination of target size while rotational dependence leads to robust behavior in target orientation identification.

In this study we do not examine the detection theory aspect of ATR. Murenzi, Kaplan, and Mujica [2] have applied the 2D CWT to detection and found it effective at making targets stand out from their backgrounds.

There are a number of distinct approaches to recognition. The most often utilized techniques are model-based target recognition, pattern recognition, and artificial neural networks. Chan and Nasrabadi [3, 4] have recently proposed an algorithm that combines attributes of both pattern recognition and neural networks techniques.

The artificial neural network is an experiential learning-by-example method that loosely models some of the aspects of the human cognitive system, the most flexible and robust target-recognition system for imagery. The goal of neural network approaches is to develop a system that reproduces this flexibility and robustness.

The main characteristics of model-based target recognition are (1) matching of processed data on the basis of multiple localized features such as the gun, turret, or body of a tank; and (2) matching of processed data to predictions based on hypotheses concerning the target type, pose, and range to increase computational speed and efficiency.

Pattern recognition, the method used in this paper, is the most mature approach used for ATR applications. In this case, target signatures are represented in feature vectors and recognition involves selecting the best match between the feature vector of the target and feature vector templates in a database. The matching criteria may be ad hoc (e.g. mean-square differences between target vectors and database templates), or statistical (such as the nearest neighbors or K-means algorithm which assumes that k reasonable clusters can be defined for

a given data set by minimizing a distance measure between the data and the *centroids* of the clusters) [5].

In all cases large amounts of data are necessary to sufficiently train the system. Currently we are using data from the TRIM2 database which is relatively small. Our investigation used about 250 images, while Chan and Nasrabadi [3] performed their algorithms on the SIG and ROI databases, which together contain over 16000 images. Clearly the data we have cannot adequately capture the range of variations in FLIR sensor data which restricts our ability to refine the algorithm enough to recognize poor quality images.

Fig 1.1 outlines our basic approach to classifying targets in our database (TRIM2). After receiving an image we ignore the detection aspect and manually extract our targets of interest. The image is then resized and normalized to allow meaningful comparison in the classification part. Next, the image is transformed to feature space using the space energy density followed by a minimum distance classifier algorithm to identify the target.

This thesis begins with Chapters 1 and 2 which include an introduction to wavelets and the continuous wavelet transform. The energy density and the motivation behind its use are discussed in Chapter 3. A brief examination of classifiers is given in Chapter 4, followed by an explanation of how the experiment was conducted. The results and conclusions are then presented at the end of the thesis.

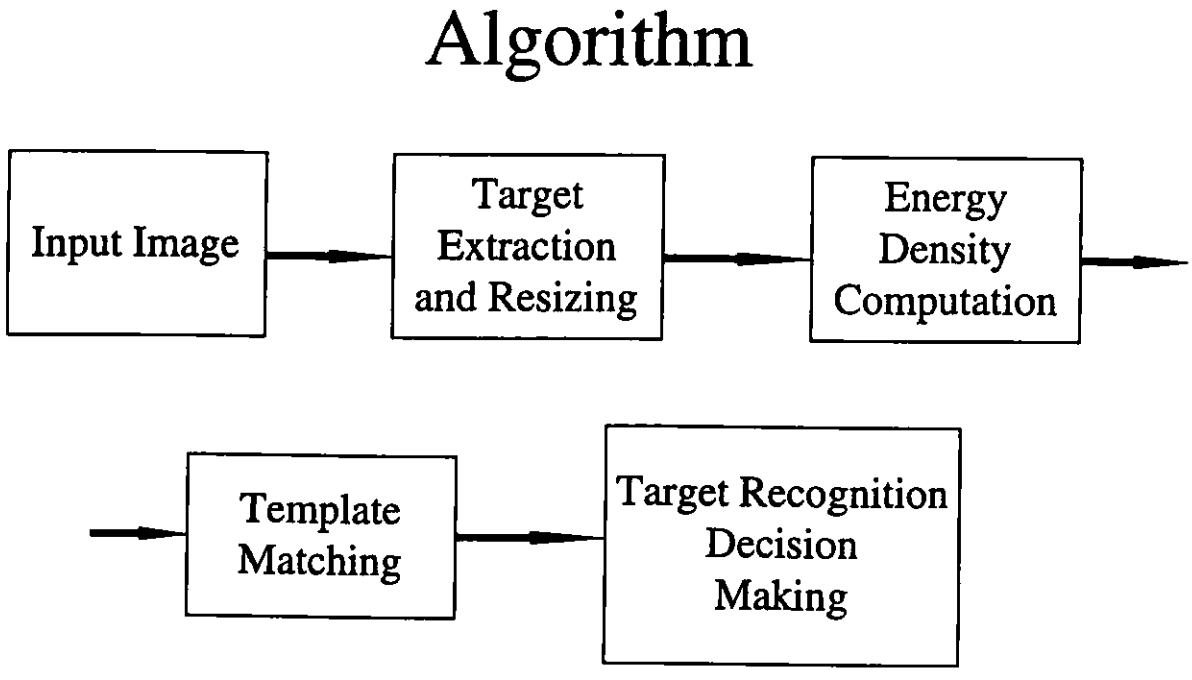

Figure 1.1: Strategy for target recognition

Chapter 2
The One Dimensional Continuous Wavelet Transform

2.1 Time-Scale Representations

The 1D CWT is often used for the analysis of signals that depend on time, such as speech signals. These signals sometimes contain physically significant transient components and cover a wide range of frequencies. In addition, there is frequently a direct correlation between the characteristic frequency and the time duration of that segment. For example, low frequencies tend to last a long interval, whereas high frequencies typically occur briefly. Human speech signals are typical in this respect, wherein vowels have a relatively low mean frequency and last quite long. On the other hand, consonants contain a wide spectrum and are often short. As a result, it is useful in signal analysis to be able to simultaneously obtain information about the frequency and time. The Fourier transform

$$\hat{f}(w) = \frac{1}{2\pi} \int_R e^{-iwt} f(t) dt \tag{2.1}$$

has been the traditional tool for analyzing signals. However, it has a shortfall, in that it gives a purely frequency domain representation which omits invaluable explicit time localization information.

Time-scale (TS) representations get around this problem by using two parameters: a which refers to scale and b which indicates the position in the signal.

Thus a general TS transform of a signal x will take the form

$$x(t) \leftrightarrow c(a,b) = \int_R \overline{\psi_{ab}}(t)x(t)dt, \tag{2.2}$$

where ψ_{ab} is the analyzing function and $\overline{\psi_{ab}}$ is its complex conjugate.

2.2 Wavelets and the Continuous Wavelet Transform

The wavelet transform is a TS transform whose analyzing function is of the form

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}). \tag{2.3}$$

The effect of a on the function ψ is a dilation or contraction while b causes a translation of ψ. The parameter a moves the wavelet around in the positive half plane of the frequency. Thus, combining equations (2.2) and (2.3), one obtains the basic formula for the 1D CWT:

$$W_x(a,b) = \frac{1}{\sqrt{a}}\int_R \overline{\psi}(\frac{t-b}{a})x(t)dt. \tag{2.4}$$

Comparing equations (2.4) and (2.1) we see that the CWT is a projection of the signal onto the family $\{\psi_{ab}, a > 0, b \in R\}$ generated from the single function ψ by translations and dilations:

$$W_x(a,b) = \langle \psi_{ab}|x\rangle \tag{2.5}$$

The property $\langle \psi_{ab} | x \rangle = \langle \widehat{\psi_{ab}} | \hat{x} \rangle$ means that the CWT can also be written as

$$W_x(a,b) = \sqrt{a} \int_{-\infty}^{\infty} \hat{\psi}(aw) \hat{x}(w) e^{ibw} dw, \tag{2.6}$$

where $\hat{\psi}$ and $\hat{x}$ denote the Fourier transforms of ψ and x, respectively. Hence, the transform $W_x(a,b)$ is in the time-scale half-plane $R_+^2 = a > 0, b \in R$.

Three properties must be satisfied: the analyzing wavelet ψ and its Fourier transform $\hat{\psi}(w)$ must be square integrable and should be well localized and ψ must be admissible, i.e.,

$$c_\psi = 2\pi \int_{-\infty}^{\infty} |\hat{\psi}(w)|^2 \frac{dw}{|w|} < \infty. \tag{2.7}$$

If ψ is regular enough ($\psi \in L^1(R) \cap L^2(R)$ suffices), then the admissibility condition simply means that the wavelet has zero mean, i.e.,

$$\hat{\psi}(0) = 0 \Leftrightarrow \int dt \psi(t) = 0. \tag{2.8}$$

The following are one dimensional versions of wavelets we use in our analysis:

1. The *Mexican hat* wavelet

$$\psi_H(t) = (1 - t^2) e^{-t^2/2}, \tag{2.9}$$

is the second derivative of the Gaussian. In the frequency domain,

$$\hat{\psi}_H(w) = w^2 e^{-w^2/2}. \tag{2.10}$$

Fig. 2.1 illustrates the localization property in both time and frequency as well as the oscillating/zero mean in the time domain.

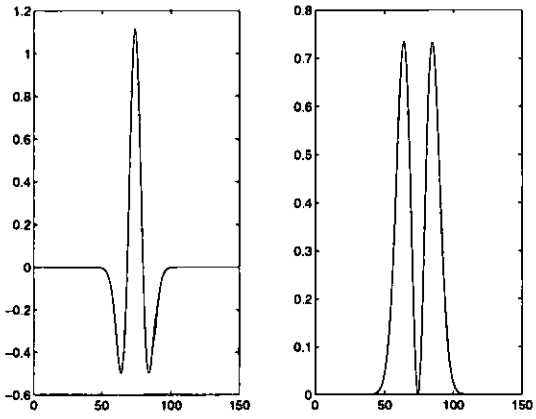

Figure 2.1: The 1D Mexican hat wavelet and its Fourier transform.

2. The *Morlet* wavelet

$$\psi_M(t) = (e^{iw_0 t} - e^{w_0^2/2})e^{-t^2/2\sigma_0^2}, \tag{2.11}$$

corresponds to a modulated Gaussian and its Fourier Transform is given by

$$\hat{\psi}_M(w) = e^{-[(w-w_0)\sigma_0]^2/2} - e^{-(w\sigma)^2/2}e^{-(w_0\sigma)^2/2}. \tag{2.12}$$

As stated earlier, wavelets must be localized in time and space. If it is assumed that ψ has "essential" supports of width T centered around 0 while $\hat{\psi}$ has an essential support of width F centered around Ω_0, then the transformed wavelets ψ_{ab} and $\hat{\psi}_{ab}$ have an essential support of width aT around b and F/a around Ω_0/a, respectively. Since the product of the two widths remains constant, a decrease in width in the time domain corresponds to an increase in width in the frequency domain. This feature is consistent with the Fourier uncertainty

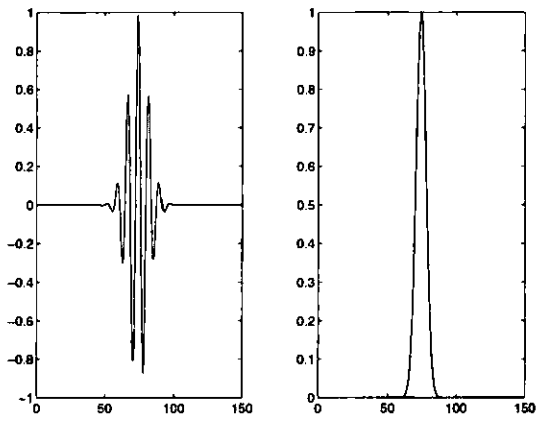

Figure 2.2: The real part of the 1D Morlet wavelet and its Fourier transform.

principle and is illustrated in Fig. 2.3. Note also that the uncertainty principle implies that there is a lower limit to the product.

Thus if $a \gg 1$, ψ_{ab} is a wide window and $\hat{\psi}_{ab}$ is highly localized around a small frequency. Interpreting this with the example of human speech, we see that a low frequency lasts for a relatively long duration. This transform will be most sensitive to low frequencies. Similarly, if $a \ll 1$, then ψ_{ab} is a narrow window and $\hat{\psi}_{ab}$ is wide and centered around a high frequency Ω_0/a. Thus the transform has a good localization capability and is sensitive to high frequencies. The CWT is a convolution. It is well known that a convolution along with the localization properties and the zero mean condition of the wavelet yield a process that performs *local filtering* both in time and frequency. The wavelet transform $W_x(a,b)$, is nonnegligible only when the wavelet "matches" components of the signal, i.e., it filters the part of the signal that lives around the time b and the scale a.

The analysis works at constant relative bandwidth ($\Delta w/w = constant$), so it has better resolution (in time) at high frequencies, i.e., small scales, which makes it an ideal tool for detecting discontinuities in the signal [6].

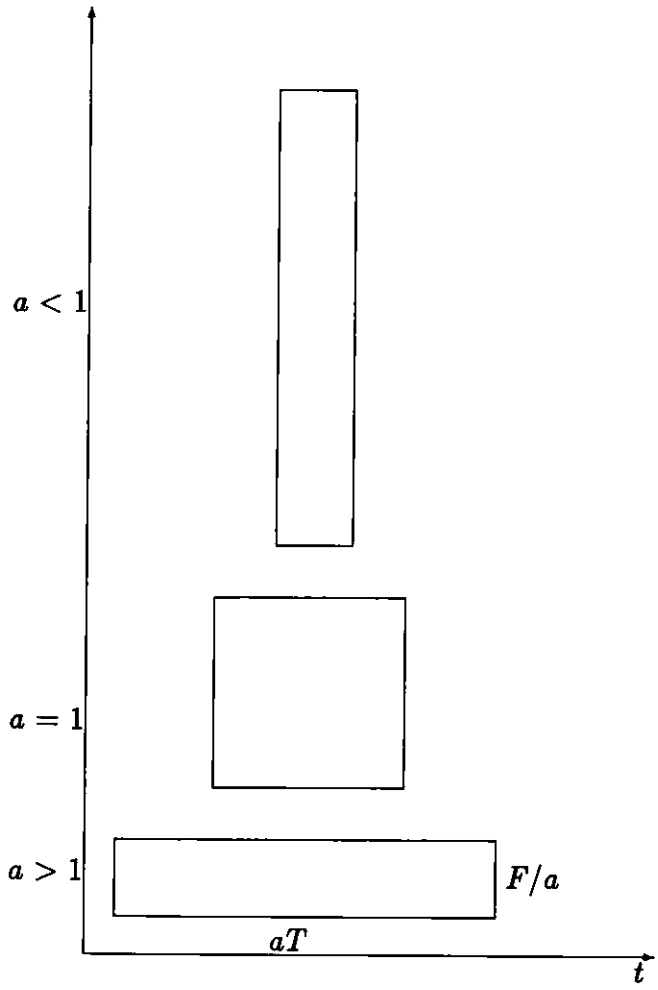

Figure 2.3: Support properties of ψ_{ab} and $\hat{\psi}_{ab}$.

Chapter 3

The Two Dimensional Continuous Wavelet Transform

3.1 Images

The 2D CWT provides a representation of an image over scaled, translated, and *rotated* versions of a single mother basis function.

We define an image as a two dimensional signal of finite energy that is a complex-valued square-integrable function $s \in L^2(R^2, d^2\vec{x})$, i.e.,

$$\|s\|^2 = \int_{R^2} d^2\vec{x}|s(\vec{x})|^2 < \infty. \tag{3.1}$$

A black and white image can be represented by a bounded non-negative function:

$$0 \leq s(\vec{x}) \leq M < \infty, \forall \vec{x} \in R^2.$$

Typically the values of $s(\vec{x})$ are quantized to a finite level of gray scale, wherein the discrete values of $s(\vec{x})$ correspond to the gray level of each pixel.

3.2 Wavelets

The Fourier transform of a signal s can be represented as

$$\hat{s}(\vec{k}) = \frac{1}{2\pi} \int_{R^2} d^2\vec{k} e^{-i\vec{k}\cdot\vec{x}} s(\vec{x}),$$

where $\vec{k} \in R^2$ is the spatial frequency and $\vec{k} \cdot \vec{x} = k_1 x_1 + k_2 x_2$ is the inner product. Note that the Fourier transform is unitary, i.e.,

$$\hat{s} \in L^2(R, d^2\vec{k}),$$

and

$$\|\hat{s}\|^2 = \|s\|^2.$$

A wavelet is a complex-valued function $\psi \in L^2(R^2, d^2\vec{x})$ that satisfies the following conditions:

1. $\hat{\psi}(\vec{k})$ is square integrable ($\hat{\psi} \in L^2(R, d^2\vec{k})$);

2. ψ is an admissible vector, i.e., the integral

$$c_\psi = (2\pi)^2 \int |\hat{\psi}(\vec{k})|^2 \frac{d^2\vec{k}}{|\vec{k}|^2} < \infty \tag{3.2}$$

converges which implies

$$\hat{\psi}(\vec{0}) = 0,$$

which in turn implies the zero mean condition

$$\int \psi(\vec{x}) d^2\vec{x} = 0.$$

The zero mean condition implies that ψ must be oscillating.

All operations that will be applied to a signal s can be obtained by combining three elementary transformations of the plane, namely translations, dilations

and rotations, which are represented by the following unitary operators in the space $L^2(R^2, d^2\vec{x})$ of signals:

1. translation: $(T^{\vec{b}}s)(\vec{x}) = s(\vec{x} - \vec{b}), \vec{b} \in R^2$
2. dilation: $(D^a s)(\vec{x}) = \frac{1}{a}s(\frac{\vec{x}}{a}), a > 0$
3. rotation: $(R^\theta s)(\vec{x}) = s(r_{-\theta})(\vec{x}), \theta \in [0, 2\pi)$ where $\vec{b}$ is the displacement parameter, $a > 0$, the dilation parameter, θ the rotation angle, and the rotation matrix $r_\theta \epsilon SO(2)$ which operates on $\vec{x} = (x, y)$:

$$r_\theta(\vec{x}) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta), 0 \leq \theta < 2\pi.$$

The translation, dilation and rotation operators generate the two-dimensional Euclidean group with dilations (the similitude group SIM(2) of R^2, denoted here by G). $\Omega(a, \theta, \vec{b}) = T^{\vec{b}}D^a R^\theta$ is a unitary irreducible representation of G in $L^2(R^2, d^2\vec{x})$. $\Omega(a, \theta, \vec{b}) = T^{\vec{b}}D^a R^\theta$ operates on a given function $s \in L^2(R^2, d^2\vec{x})$ as:

$$(\Omega(a, \theta, \vec{b})s)(\vec{x}) = s_{a,\theta,\vec{b}}(\vec{x}) \equiv \frac{1}{a}s(r_{-\theta}(\frac{\vec{x} - \vec{b}}{a})). \tag{3.3}$$

Using the following properties of the Fourier transform

1. $\mathcal{F}(s(\vec{x} - \vec{b})) = e^{-i\vec{b}\cdot\vec{k}}\hat{s}(\vec{k})$
2. $\mathcal{F}(s(\frac{\vec{x}}{a})) = a^2\hat{s}(a\vec{k})$
3. $\mathcal{F}(s(r_{-\theta}(\vec{x}))) = \hat{s}(r_{-\theta}(\vec{k}))$

the equivalent result in the space of Fourier transforms can be shown to be

$$\widehat{s_{a,\theta,\vec{b}}}(\vec{k}) = ae^{-i\vec{b}\cdot\vec{k}}\hat{s}(ar_{-\theta}(\vec{k})).$$

Note that if the function s is rotationally invariant, the index θ can be omitted:

$$s_{a,\vec{b}}(\vec{x}) = \frac{1}{a}s(\frac{\vec{x}-\vec{b}}{a}).$$

The three unitary operators $T^{\vec{b}}$, D^a, R^θ and their combination $\Omega(a,\theta,\vec{b})$ preserve the admissibility condition. Clearly since $\Omega(a,\theta,\vec{b})$ does not affect any function that satisfies conditions (1) and (2) of the CWT, any function $\psi_{a,\theta,\vec{b}} = \Omega(a,\theta,\vec{b})\psi$ obtained from the wavelet ψ by a translation, dilation, or rotation is also a wavelet. Thus the given mother wavelet ψ generates an infinite family of wavelets $\{\psi_{a,\theta,\vec{b}}\}$ indexed by elements $a > 0$, $\theta \in [0, 2\pi)$, $\vec{b} \in R^2$.

3.3 The Continuous Wavelet Transform

Let $s \in L^2(R^2, d^2\vec{x})$ be an image. Its *continuous wavelet transform* (with respect to the fixed wavelet ψ) $S \equiv W_\psi s$, is the scalar product of s and the transformed wavelet $\{\psi_{a,\theta,\vec{b}}\}$ considered to be a function of $(a,\theta,\vec{b}) \in G$:

$$\begin{aligned} S(a,\theta,\vec{b}) &= \langle\psi_{a,\theta,\vec{b}}|s\rangle, && (3.4)\\ &= \frac{1}{a}\int d^2\vec{x}\overline{\psi}(r_{-\theta}(\frac{\vec{x}-\vec{b}}{a}))s(\vec{x}) && (3.5)\\ &= a\int d^2\vec{k}e^{i\vec{b}\cdot\vec{k}}\overline{\hat{\psi}}(ar_{-\theta}(\vec{k}))\hat{s}(\vec{k}). && (3.6) \end{aligned}$$

The wavelet may be chosen to have a few vanishing moments which determine the capacity of the wavelet to detect or ignore singularities of various order derivatives in the image.

If ψ has n vanishing moments, i.e.,

$$\int d^2\vec{x}x^\alpha y^\beta\psi(\vec{x}) = 0, 1 \le \alpha+\beta \le n, \qquad (3.7)$$

then the wavelet transform (WT)W_ψ is blind to polynomials of degree up to n. This feature directly affects the ability of the WT to detect singularities in derivatives of order less than n.

Some important properties of the CWT $W_\psi : s \in L^2(R^2, d^2\vec{x}) \mapsto S \in L^2(G, dg)$, $dg \equiv a^{-3}dad\theta d^2\vec{b}$ defined by $[W_\psi s](a, \theta, \vec{b}) = c_\psi^{-1/2}\langle \psi_{a,\theta,\vec{b}} | s \rangle$ are:

1. W_ψ is *linear* in the signal s;

2. W_ψ is *covariant* under translations, dilations and rotations which means that the correspondence $W_\psi : s(\vec{x}) \mapsto S(a, \theta, \vec{b})$ implies the following:

$$W_\psi : s(\vec{x} - \vec{b}_0) \mapsto S(a, \theta, \vec{b} - \vec{b}_0),$$

$$W_\psi : \frac{1}{a_0} s(\frac{\vec{x}}{a_0}) \mapsto S(\frac{a}{a_0}, \theta, \frac{\vec{b}}{a_0}), \tag{3.8}$$

$$W_\psi : s(r_{\theta_0}(\vec{x})) \mapsto S(a, \theta - \theta_0, r_{-\theta_0}(\vec{b})).$$

3. W_ψ *conserves energy*:

$$c_\psi^{-1} \int\int\int_G \frac{da}{a^3} d\theta d^2\vec{b} |S(a, \theta, \vec{b})|^2 = \int d^2\vec{x} |s(\vec{x})|^2, \tag{3.9}$$

i.e., it is a linear isometry from the space of signals $L^2(R^2, d^2\vec{x})$ into the space of transforms $\mathcal{H}_\psi$, a closed subspace of $L^2(G, dg)$, where $dg \equiv a^{-3}dad\theta d^2\vec{b}$.
Since it is an isometry, the map W_ψ is invertible on its range $\mathcal{H}_\psi$, and the inverse transformation is simply the adjoint of W_ψ. Thus the image $s(\vec{x})$ can be reconstructed from its wavelet transform $S(a, \theta, \vec{b})$ by the formula

$$s(\vec{x}) = c_\psi^{-1/2} \int\int\int_G \frac{da}{a^3} d\theta d^2\vec{b} \psi_{a,\theta,\vec{b}}(\vec{x}) S(a, \theta, \vec{b}). \tag{3.10}$$

The 2D CWT provides a decomposition of the signal in terms of the analyzing wavelets $\psi_{a,\theta,\vec{b}}$, with coefficients $S(a,\theta,\vec{b})$.

3.4 The Two Basic Representations

The basic problems of the 2D CWT involve computation and visualization. $S(a,\theta,\vec{b})$ is a function of four variables: two position $\vec{b} \in R^2$, and the pair $(a,\theta) \in R_*^+ \times [0,2\pi) \simeq R_*^2$. Thus the CWT w_ψ unfolds the signal from two to four dimensions. Although this feature gives the CWT efficiency in decoupling singularities, it increases the data load. By fixing some variables this problem can be overcome. Six different visualization possibilities exist. If the parameter space is viewed as a phase space [7], then two natural representations of the CWT based on the use of two-dimensional sections of the parameter space can be found [8, 9]:

1. the *position representation*, where a and θ are fixed and the CWT is considered as a function of position $\vec{b}$ alone;

2. the *scale-angle representation*, which for fixed $\vec{b}$, the CWT is considered as a function of scale and angle (a,θ), i.e., of spatial frequency.

The position representation is the standard representation and is used in this study; it is useful for the detection of position, shape and contours of objects. The scale-angle representation is useful in detection for angular selection and scaling.

3.5 Analyzing Wavelets

Before the CWT is applied, an analyzing wavelet must be chosen. There are standard wavelets and wavelets that are adapted to the problem at hand. In

the early stages of the analysis, it is instructive to use the standard wavelets, and advance to the adapted wavelets once a feel for the problem is obtained. In our analysis, we use the 2D Mexican hat and Morlet wavelets.

3.5.1 The 2D Mexican Hat Wavelet

In its isotropic version, 2D Mexican hat wavelet corresponds to the Laplacian of the Gaussian (Fig. 3.1).

$$\psi_H(\vec{x}) = (2 - |\vec{x}|^2) \exp(-\frac{1}{2}|\vec{x}|^2). \tag{3.11}$$

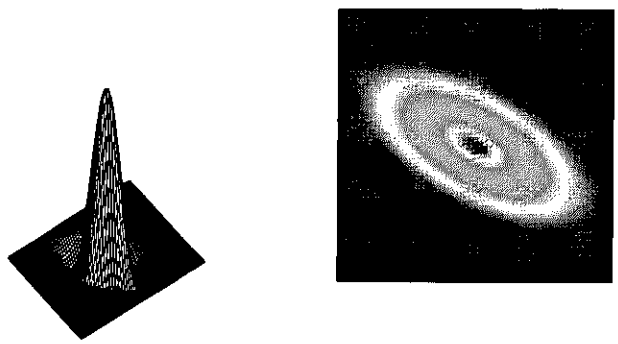

Figure 3.1: The 2D anisotropic Mexican hat wavelet ψ_H with $\epsilon = 5, a = 1, \theta = \pi/3$: (left) in position space; (right) in spatial frequency space.

The wavelet is real, rotationally invariant and has vanishing moments of order 0 and 1 which enable it to detect singularities in the first derivative of the signal. An anisotropic version exists,

$$\psi_H(\vec{x}) = (2 - |A\vec{x}|^2) \exp(-\frac{1}{2}|A\vec{x}|^2), \tag{3.12}$$

where

$$A = \begin{bmatrix} \epsilon^{-1/2} & 0 \\ 0 & 1 \end{bmatrix}, \epsilon \geq 1 \tag{3.13}$$

is the anisotropy matrix. This version however is not very useful in practice since it is still a second order operator and detects singularities in all directions. It is not a directional wavelet and cannot detect oriented texture information. It will thus only be used for pointwise analysis.

3.5.2 The 2D Morlet Wavelet

The 2D Morlet wavelet 3.2 is given by

$$\psi_M(\vec{x}) = \exp(i\vec{k}_0 \cdot \vec{x}) \exp(-\frac{1}{2}|A\vec{x}|^2) + corr. \tag{3.14}$$

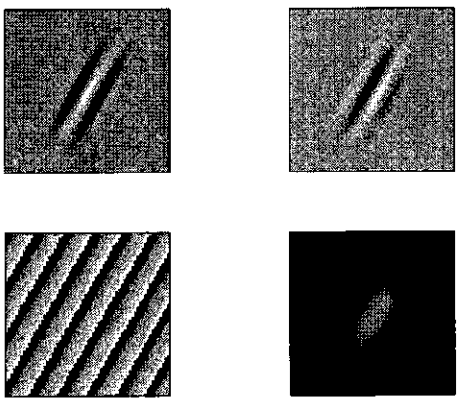

Figure 3.2: The 2D anisotropic Morlet wavelet ψ_M with $\epsilon = 5$, $\vec{k}_0 = (0, 3.2)$, $a = 5$,and $\theta = \pi/3$ in position space: (top) real and imaginary parts; (bottom) phase and modulus.

The correction term is $exp[-\frac{1}{2}(|A\vec{x}|^2 + |\vec{k}_0|^2)]$. Without it $\psi_M(\vec{x})$ is the Gabor function, which is not a wavelet since it does not satisfy the zero mean condition. The correction term overcomes this by forcing $\hat{\psi}(\vec{0}) = 0$. Although it

is considered numerically negligible for $|\vec{k}_0| \geq 5.6$, and hence is usually dropped, we have found in our analysis that ignoring it destroys its wavelet properties. A is the anisotropy matrix given in equation (3.13) and the parameter $\vec{k}_0$ is the wave vector. The phase of the wavelet is constant in the direction perpendicular to $\vec{k}_0$, which is very useful because the wavelet smoothes the signal in all directions and detects singularities in the direction perpendicular to $\vec{k}_0$. Angular selectivity increases with $|\vec{k}_0|$ and even more so if an anisotropy is introduced by taking $\epsilon > 1$.

3.5.3 Directional Wavelets

To detect oriented features such as edges and texture, a directionally sensitive wavelet is needed. A wavelet ψ is *directional* if the effective support of its Fourier transform $\hat{\psi}$ is contained in a convex cone in spatial frequency space $\{\vec{k}\}$, with apex at the origin, or in a finite union of disjoint such cones (ψ is then referred to as *multidirectional*).

The support of the anisotropic Mexican hat wavelet $\hat{\psi}$ in spatial frequency

$$\hat{\psi}_H(\vec{k}) = (\epsilon k_x^2 + k_y^2) \exp -\frac{1}{2}(\epsilon k_x^2 + k_y^2), \tag{3.15}$$

is centered at the origin irrespective of the size of the anisotropy as is shown in Fig. 3.1. Detailed tests confirm it to be poor in selecting directions [8]. The Fourier transform ψ of the Morlet wavelet

$$\hat{\psi}_M(\vec{k}) = \sqrt{\epsilon} \exp(-\frac{1}{2}[\epsilon k_x^2 + (k_y - k_0)^2]), \tag{3.16}$$

is consistent with the definition of a directional wavelet. Its effective support is centered at $\vec{k}_0$ and is contained in a convex cone that becomes narrower as ϵ

increases. ψ_M with $\epsilon = 5$ and rotated by $\theta = \pi/3$ is shown in Fig. 3.3.

3.6 Performance Evaluation of the 2D CWT

The scale and angular selectivity of a wavelet lies in a quantative knowledge of the properties of the wavelet; hence the wavelet must be calibrated. Calibration can be accomplished by taking the wavelet transform of particular standard signals. There are three instructive tests:

- *Point signal.* For a snapshot of the wavelet itself, one takes as the signal a delta function i.e., one evaluates the response of the filter

$$\langle \psi_{a,\theta,\vec{b}} | \delta \rangle = \frac{1}{a} \overline{\psi} \left(\frac{1}{a} r_{-\theta}(-\vec{b}) \right), \tag{3.17}$$

  which yields the effective support of ψ.

- *Benchmark signals.* For testing particular properties of the wavelet such as its ability to detect discontinuities or its angular selectivity in detecting a particular direction, appropriate 'benchmark' signals are used.

Figure 3.3: The Morlet wavelet ψ_M in spatial frequency space with $\epsilon = 5$, $\vec{k}_0 = (0, 3.2)$, $a = 2$, and $\theta = \pi/3$.

- *Reproducing kernel.* The wavelet transform of the wavelet is the reproducing kernel K, which measures the *correlation length* in each variable $a, \theta, \vec{b}$:

$$K(a,\theta,\vec{b}|1,0,\vec{0}) = \langle \psi_{a,\theta,\vec{b}} | \psi \rangle = \frac{1}{a} \int d^2\vec{x} \overline{\psi} \left(\frac{1}{a} r_{-\theta}(\vec{x} - \vec{b}) \right) \psi(\vec{x}). \qquad (3.18)$$

An analysis of K gives rise to a definition of the resolving power of the wavelet ψ in each variable.

3.6.1 Calibration of a Wavelet with Benchmark Signals

As with any tool, calibration of a wavelet with respect to a given process requires the use of benchmark signals. A systematic analysis is presented in [8] for detection of a line singularity (rod), detection of a given direction (several segments of a rod at different orientations), and measurement of the opening angle of a wedge.

In all cases the response of the CWT is given as a function of the various parameters of the wavelet used, such as the wave number k_0, anisotropy ϵ and orientation, in the case of the Morlet wavelet.

A general conclusion is the Morlet wavelet is efficient for a sufficiently large value of the product $k_o\sqrt{\epsilon}$ and is able to detect the orientation of a segment with a high degree of selectivity, of the order of a few degrees. An example of its selectivity is given in [7]. Thus it is efficient also for *directional filtering.*

The Mexican hat is found to be efficient at detecting contours by detecting *discontinuities* in images. The isotropic Mexican hat is sufficient for this purpose. The effect of the wavelet transform with the Mexican hat ψ_H is the smoothing of the signal with a Gaussian and taking the Laplacian of the result. Large values of

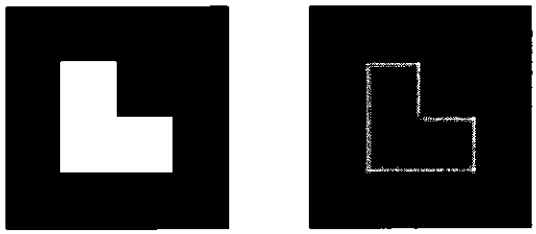

Figure 3.4: The letter "L" and its wavelet transform using the Mexican hat wavelet.

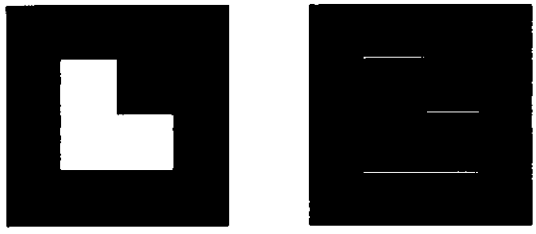

Figure 3.5: The Letter "L" and its wavelet transform using the Morlet wavelet oriented at 90°.

the amplitude appear at the location of the discontinuities, in particular, the contours of objects (which in gray scale images is a discontinuity in the luminosity). In Fig. 3.4 shows the wavelet transform of an object in the shape of the letter 'L' with the Mexican hat wavelet. For large values of a the wavelet transform sees only the object as a whole, which allows the determination of its position in the plane. For smaller values of a increasingly finer details appear and the wavelet transform vanishes both inside and outside the contour since the signal is constant there, leaving behind only the contour. Fig. 3.5 shows the directional filtering of the Morlet wavelet, in which only contours perpendicular to k_0 are highlighted.

3.6.2 The Reproducing Kernel and the Resolving Power of the Wavelet

The best way to test the correlation length of the wavelet is to analyze its reproducing kernel K. First, one can plot K in the scale-angle representation, i.e., in polar coordinates (a, θ). From the plot, the angular width $\Delta\theta$ of K and its scale range $\Delta a = a_{\max}/a_{\min}$ can be obtained, where $[a_{\max}, /a_{\min}]$ is the effective support of K in the variable a. Let the wavelet ψ have its effective support in spatial frequency in a cone of aperture $\Delta\phi = \phi_2 - \phi_1$, between scales ρ_1 and ρ_2. The effective support of the reproducing kernel is given by $a_{\min} = \rho_1/\rho_2$ and $a_{\max} = \rho_2/\rho_1$ for the scale variable and $-\Delta\phi \leq \theta \leq \Delta\phi$ for the angular variable. The resolving power can be defined in terms of the following parameters of K:

1. scale width or scale resolving power, $\Delta\rho = \rho_2/\rho_1 = \sqrt{\Delta a}$; and

2. angular width or angular resolving power, $\Delta\phi = \frac{1}{2}\Delta\theta$,

which is useful for determining the minimal sampling grid required for the numerical evaluation of the reconstruction integral (equation 3.10).

3.6.3 Scale and Angular Resolving Powers: Numerical Evaluation

Given a directional wavelet ψ, its effective support in spatial frequency space can be characterized. A wavelet ψ is located at $\vec{k}_0$ if

$$\vec{k}_0 = \int d^2\vec{k}\vec{k}|\hat{\psi}(\vec{k})|^2. \tag{3.19}$$

Similarly, its width in the direction ξ is given by $2w_\xi$, where

$$w_\xi = \frac{1}{\|\hat{\psi}\|}\left[\int d^2\vec{k}|(\vec{k}-\vec{k}_0)\cdot\vec{\xi}|^2|\hat{\psi}(\vec{k})|^2\right]^{1/2}. \tag{3.20}$$

If after a suitable rotation the supporting cone is vertical, i.e., $\vec{k}_0 = (0, k_0)$, then the widths of ψ in the horizontal and vertical directions are given by $2w_x$ and $2w_y$ respectively, as shown in Fig. 3.6., where

$$w_x = \frac{1}{\|\hat{\psi}\|}\left[\int d^2\vec{k}k_x^2|\hat{\psi}(\vec{k})|^2\right]^{1/2}, \tag{3.21}$$

and

$$w_y = \frac{1}{\|\hat{\psi}\|}\left[\int d^2\vec{k}(k_y-k_0)^2|\hat{\psi}(\vec{k})|^2\right]^{1/2}. \tag{3.22}$$

Antoine and Murenzi [7] show that the scale resolving power (SRP) is

$$SRP(\psi) = \frac{k_0 + w_y}{k_0 - w_y}, \tag{3.23}$$

and the angular resolving power (ARP) is

$$ARP(\psi) = 2\cot^{-1}\frac{\sqrt{k_0^2 - w_y^2}}{w_x}. \tag{3.24}$$

The SRP and the ARP give the effective support of a wavelet in Fourier space, which allows choice of scales and angles for a directional wavelet in such a way that together the individual wavelets fill as much of the frequency space as possible with no overlap of the respective supports (see Fig. 3.7).

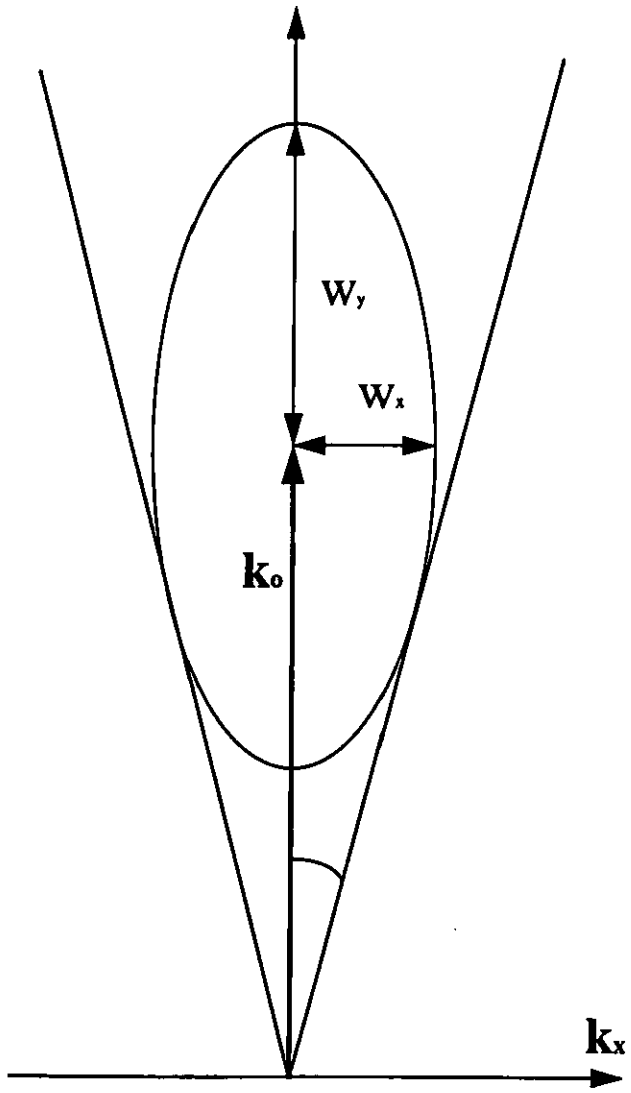

Figure 3.6: Evaluation of the resolving power of a wavelet.

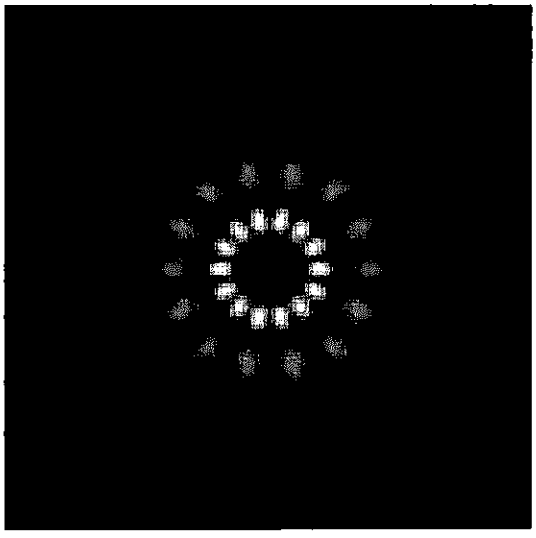

Figure 3.7: Rotated and scaled Morlet wavelets tiled together in spatial frequency space.

Chapter 4

Continuous Wavelet Transform-Based Automatic Target Recognition

4.1 Introduction

ATR requires the ability to extract the essential features of objects in cluttered environments. In this work the objects are tanks imaged with forward-looking infrared (FLIR) sensors. FLIR images display a large amount of variability, particularly with the low resolution images we are using. Aside from the usual variations that come from different aspect and elevation angles, there are differences that depend on the temperature of the target and its surrounding clutter, i.e., the target can appear hotter or colder than its environment, which results in different gray-scale images of the same target. Most of the time the target does not appear in ideal conditions, which makes it difficult to estimate the sensor output probability density function (PDF). It is therefore better to use an algorithm that extracts the features of interest (such as structural, spatial, and frequency features) and performs classification based on those features. Der and Chellappa [1] give a thorough discussion on these issues. Multiscale techniques such as the multidimensional continuous wavelet transform (MCWT) are highly desirable because they can extract and normalize the unknown scale and orientation of the target.

To extract features such as the orientation, spatial extent, shape, and symmetry of a target, it is preferable to work in the feature space of position, scale, and orientation, rather than in the image space. As explained in Chapter 2, the 2D CWT decomposes the image onto a set of scaled, translated and rotated "daughter" wavelets $\psi_{a,\theta,\vec{b}}$ of a single mother wavelet ψ. The scale dependence allows sensitivity to variations in sensor resolution in addition to determination of target size.

Integration over all the parameters of the squared modulus of the CWT gives the total energy in the signal. Therefore, integration on a subset of the parameters gives an energy density in the remaining variables, which yields in two dimensions a total of six energy densities, the CWT features. In this study we investigate the position energy density, which is an integration over scales and angles. Fig. 1.1 shows the five stages of a modification of an algorithm described in [4]. We start by extracting the target from most of its background using a border-clipping rectangular window. The extracted area is enlarged to a fixed size and feature extraction is accomplished by computing the position energy density. The classification stage involves template matching and a distance calculation for the decision.

4.2 Energy Density

For the image $s \in L^2(R^2, d^2\vec{x})$, $|s(\vec{x})|^2$ and $|\widehat{s(\vec{k})}|^2$ are energy densities. Since the wavelet transform W_ψ conserves the energy of a signal we have the following:

$$\int d^2\vec{x}|s(\vec{x})|^2 = \int d^2\vec{k}|\widehat{s(\vec{k})}|^2 = c_\psi^{-1} \int\int\int_G \frac{da}{a^3} d\theta d^2\vec{b}|S(a,\theta,\vec{b})|^2, \qquad (4.1)$$

where $S(a,\theta,\vec{b}) \in (L^2(R_+^* \times ([0,2\pi]) \times R^2, \frac{da}{a^3}d\theta d^2\vec{b})) = W_\psi s(x)$ is the wavelet transform of the signal s. Thus we can interpret the $|S(a,\theta,\vec{b})|^2$ as an energy density of the signal in position, scale and orientation variables. As mentioned above, we have six energy densities:

- Space (range and aspect) energy density:

$$E_{12}(b_x,b_y) = \int_0^\infty \frac{da}{a^3}\int_0^{2\pi} d\theta |S(a,\theta,b_x,b_y)|^2 \tag{4.2}$$

- Scale-angle energy density:

$$E_{34}(a,\theta) = \int_{-\infty}^{+\infty} db_x \int_{-\infty}^{+\infty} db_y |S(a,\theta,b_x,b_y)|^2 \tag{4.3}$$

- Anisotropy and aspect energy density:

$$E_{13}(\theta,\alpha) = \int_0^\infty \frac{da}{a^3}\int_0^\infty d|\vec{b}||\vec{b}||S(a,\theta,|\vec{b}|,\alpha)|^2 \tag{4.4}$$

- Scale-range energy density:

$$E_{13}(a,|\vec{b}|) = \int_0^{2\pi}\int_0^{2\pi} d\alpha |S(a,\theta,|\vec{b}|,\alpha)|^2 \tag{4.5}$$

- Anisotropy-range energy density:

$$E_{23}(\theta,|\vec{b}|) = \int_0^\infty \frac{da}{a^3}\int_0^{2\pi} d\alpha |S(a,\theta,|\vec{b}|,\alpha)|^2 \tag{4.6}$$

- Scale-aspect energy density:

$$E_{14}(a,\alpha) = \int_0^{2\pi} \int_0^{\infty} d\theta d|\vec{b}||\vec{b}||S(a,\theta,|\vec{b}|,\alpha)|^2 \tag{4.7}$$

4.3 Target Extraction and Resizing

Normally a target is surrounded by some background information that is irrelevant for the correct recognition of the target. The clutter information creates unwanted variations in the training and testing phases. Therefore proper removal of the unwanted background is essential for achieving good recognition.

Fig. 4.1 shows a typical image with a set of targets at a specific aspect angle in the TRIM2 database. We remove the background by using rectangular windows. Although it is possible to do this automatically with target-centered images, the ground truth images supplied in the TRIM2 database have inaccurate specifications, making this difficult for us to do. As a result, target chips are extracted manually. At this time we ignore two out of four chips in each set because the borders of the target are ill defined. After background removal the extracted target is enlarged to a fixed size by interpolation, which is crucial since it provides a more uniform similarity measure in terms of dimension and it allows for comparison of targets of different ranges. The targets should be enlarged to at least the size of the largest extraction window. However, for computational convenience we choose a size that is less than that, i.e., we reduce the size of some extractions which amounts to throwing away useful data. This practice for our particular application was found to have little effect on the results. A normalization is performed so that a possible match is not rejected simply because

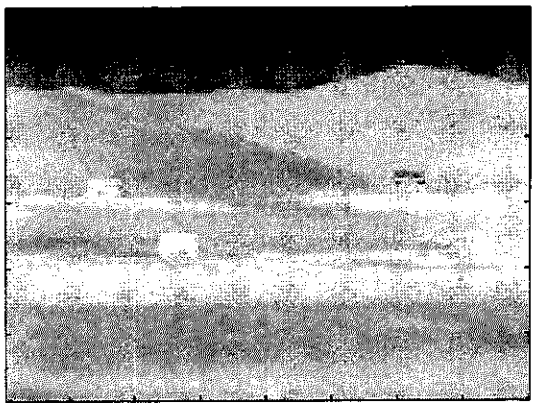

Figure 4.1: Typical target set of four at a given aspect angle in the TRIM2 database.

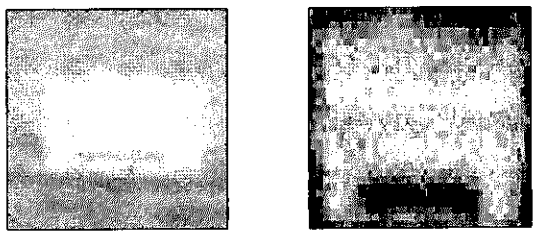

Figure 4.2: M1 tank after background removal (left). Resized and normalized M1 tank.

of differences in brightness. Fig. 4.2 shows an extracted target and what it looks like after resizing and normalizing.

4.4 Classification

4.4.1 Introduction

Classification theory in the context of mathematical pattern recognition is a mature field. Aided by tools of statistics, probability, information and automata theory, as well as other branches of applied mathematics, classification

techniques have been developed which provide a fairly sound basis upon which decision algorithms can be implemented. There is a wealth of statistical classification algorithms developed originally for detection theory in early radar and communication systems. On the other side of classification theory is an approach completely independent of statistical knowledge or assumptions. Often referred to as distribution free or nonparametric classification this approach involves the concept of a *discriminant function* and algorithms that can be described without reference to probability distributions.

As discussed Chapter 1, we are focusing on non-statistical classification techniques. We, however, do not delve into the numerous techniques that have been developed as we do not have the data to implement them (a good introductory account can be found in Andrews [11]). Moreover, we are interested in developing feature extraction techniques for input into a convolutional neural network, which will serve as our classifier.

4.4.2 Discriminant Functions

Consider the K pattern classes: $S_1, S_2, ..., S_k, ..., S_K$ with defining prototypes $\vec{k}_m^{(k)}$ where $m = 1, ..., M_K$ counts the number of prototypes within a given class. Ideally one would like a function which measures each point in the pattern or feature space and assigns to that point a value as to its degree of membership to a given class. In pattern recognition such functions are referred to as discriminant functions, which have the property that they partition the pattern or feature space into mutually exclusive regions, each region contributing to the domain of a class. The discriminant function will be defined such that for all points $\vec{x}$ within

the region describing S_k, there exists a function $g_k(\vec{x})$ such that $g_k(\vec{x}) > g_j(\vec{x})$ for all $j \neq k$. Mathematically,

$$g_k(\vec{x}) > g_j(\vec{x}) \; \forall \vec{x} \in S_k \text{ and } \forall j \neq k. \tag{4.8}$$

Thus within region S_k, the kth discriminant function will have the largest value. The surface separating region S_k and S_j is given by $g_k(\vec{x}) - g_j(\vec{x}) = 0$ which is equivalent to those points in the space which have equal discriminant functions for both class S_k and S_j. The construction and adjusting of discriminant functions is referred to as "training" or "learning". If the training is based upon statistics, parametric and certain nonparametric techniques are used. If the training is based on an assumed functional form for the discriminant function (i.e. linear, quadratic, etc.), distribution free techniques are employed. The simplest functional form for the discriminant function is known as the linear discriminant function which can be represented in scalar and vector forms as

$$g_k(\vec{x}) = w_1^{(k)} x_1 + w_2^{(k)} x_2 + .. + w_N^{(k)} x_N + w_{N+1}^{(k)} \tag{4.9}$$

or

$$g_k(\vec{x}) = \vec{w}_k^t \vec{x}. \tag{4.10}$$

Note that a scalar term $w_{N+1}^{(k)}$ has been added to the discriminant function for coordinate normalization purposes. To make equation (4.10) a valid vector multiplication, the input vector $\vec{x}$ has been augmented to become $N+1$ dimensional by $x_{n+1} = 1$.

One of the simplest classification algorithms which utilizes a linear discriminant function is the minimum distance classifier. The following simple example

is used in this study: if the average point of the prototypes defining a given class S_k be given by

$$\langle \vec{y}_k \rangle = \frac{1}{M_k} \sum_{m=1}^{M_k} \vec{y}_m^{(k)} \tag{4.11}$$

then there exist K such points in N space. Assume a Euclidean metric in the space and classifier assigned an unknown point $\vec{x}$. The decision rule then becomes

$$\vec{x} \in S_j \text{ if } d(\vec{x}, \langle \vec{y}_j \rangle) = \min_k d(\vec{x}, \langle \vec{y}_k \rangle). \tag{4.12}$$

However,

$$d^2(\vec{x}, \langle \vec{y}_k \rangle) = (\vec{x} - \langle \vec{y}_k \rangle)^t (\vec{x} - \langle \vec{y}_k \rangle) \tag{4.13}$$

$$= \vec{x}^t \vec{x} - 2\vec{x}^t \langle \vec{y}_k \rangle + \langle \vec{y}_k \rangle^t \langle \vec{y}_k \rangle. \tag{4.14}$$

According to the properties of a discriminant function we can subtract the constant $\vec{x}^t\vec{x}$ from the metric without changing the decision surfaces. Multiplication by $-\frac{1}{2}$ makes the modified distance squared function a valid discriminant:

$$g_k(\vec{x}) = \vec{x}^t \langle \vec{y}_k \rangle - \frac{1}{2} \langle \vec{y}_k \rangle^t \langle \vec{y}_k \rangle. \tag{4.15}$$

In the context of linear discriminants, the elements of $\langle \vec{y}_k \rangle$ become the linear weights and $-\frac{1}{2}\langle \vec{y}_k \rangle^t \langle \vec{y}_k \rangle$ becomes the augmenting quantity. The decision surface becomes the perpendicular bisector separating $\langle \vec{y}_j \rangle$ and $\langle \vec{y}_k \rangle$.

For a known set of prototypes $\vec{y}_m^{(k)}$ if there exist linear discriminant functions $g_1, ..., g_K$ such that $g_k(\vec{y}_m^{(k)}) > g_j(\vec{y}_m^{(k)})$ for all $m = 1, ..., M_k$ and for all $k \neq j$, then the surfaces are said to be *linearly separable*. If the set of prototypes is linearly separable, it is possible to develop algorithms to find linear

hyperplanes which properly separate the data. These algorithms are referred to as "error-correction" training procedures. Unfortunately these algorithms require large amounts of data which preclude us from using them.

4.5 Experimentation

Our input images are, as previously stated, from the FLIR database. Images in that database are separated into three different clutter backgrounds: clutters I, II, and III. So far we have only used clutters I and II. For each clutter every target chip is imaged at 21 different aspect angles with a fixed elevation angle. Each image has four targets at various distances, as shown in Fig. 4.1. In clutter I two targets are at approximately 1600 meters and the other two are at about 1350 meters. In clutter II two of the tanks are at approximately 1330 meters, and the other two are located at around 1550 meters.

4.5.1 Background Removal

We manually cut out a large 64×128 pixel window from around each of the tanks using the ground truth information in the specifications to obtain images like those shown in Fig. 4.3. We then remove most of the background from using rectangular windows to obtain a smaller sized image, say 32×47, and then resize it to 64×64 and normalize it (see Fig. 4.4). As mentioned earlier, some of the target chips are larger than 64×64, so we lose some information because we have to shrink them.

It is possible to perform the segmentation automatically from the original 480×680 pixel image given accurate ground truth information. However, because the TRIM2 data set has only approximate specifications, we choose to extract

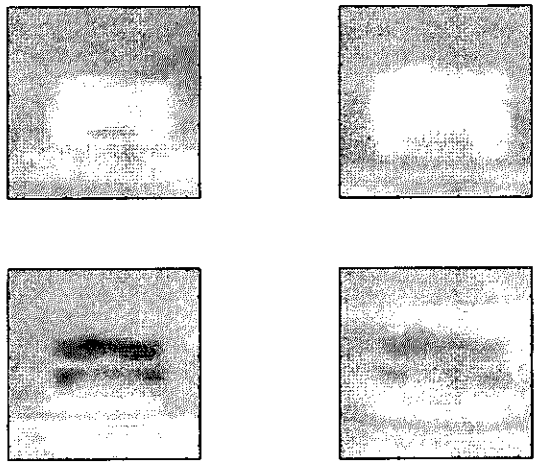

Figure 4.3: Target set of four cutout tanks at a given aspect angle in the TRIM2 database.

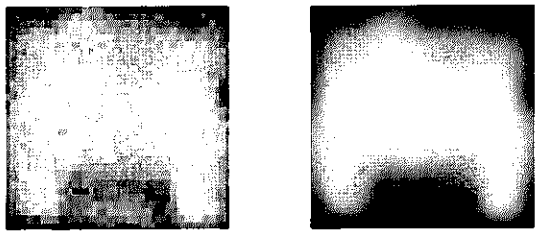

Figure 4.4: Image before and after energy density computation with Mexican hat wavelet.

the targets manually. The whole process is repeated for six target classes out of clutters I and II, and each of the 21 aspect angles.

We have chosen to ignore two out of four chips in each image because of poorly defined borders.

4.5.2 Energy Density Computation

The energy density is calculated for each 64×64 image in spatial frequency space using the CWT given in equation (3.6), enabling us to use the fast Fourier transform (FFT) which reduces the computational complexity significantly. We approximate the position energy density by taking the sum of the square of the modulus of the wavelet transform for specific scales and angles. The scales and angles are chosen according to the wavelet used.

Since we use only the isotropic version of the Mexican hat it is unnecessary to rotate the wavelet. The scales are chosen to "match" features of interest in the image such as the whole target and some of its smaller components. Due to the resolution of the images, we cannot choose arbitrarily small scales since there is a lower limit below which the wavelet begins to lose some critical features such as the zero mean condition. Additionally some of the sensor noise and clutter is spread out over the images in "packets" that are only a few pixels small, thus it is important not to choose scales of that size and hence filter out the noise.

The scales and angles for the Morlet wavelet are chosen so that rotated and scaled wavelets fill out most of the spatial frequency space of the image. As a result, the effective support of the wavelet in spatial frequency must be known. To find the effective support we must compute the scale and angular resolving powers of the the wavelet. The underlying theory is presented in Chapter 3. Fig. 3.7 shows the Morlet wavelet rotated and scaled in spatial frequency space.

4.5.3 Classification

At this stage we are only comparing tanks at the same aspect angles, i.e., if we get a tank at ten degrees from clutter II, we classify it by matching it with clutter I templates at ten degrees. Fig 4.5 shows clutter I targets at zero degrees.

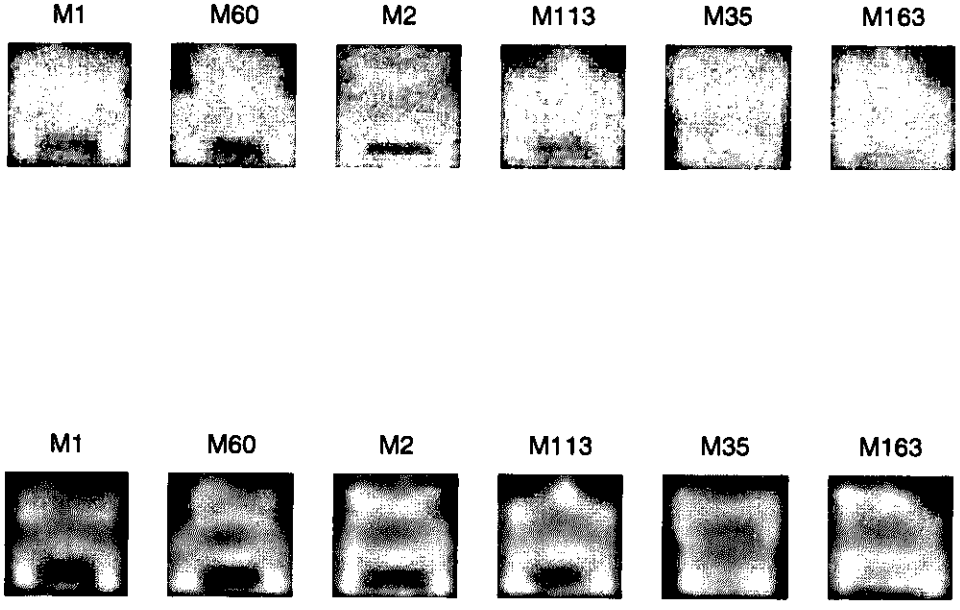

Figure 4.5: "Training" target chips at 0° from clutter 1: (top) image space target; (bottom) corresponding energy densities.

Templates are created from targets in clutter I using the formula given in equation (4.11). In this case a class is identified as a set of two prototypes at one of the 21 aspect angles for each target chip. Hence we have 21 templates for each of the six different targets. Fig. 4.6 shows a typical set of templates.

Each template is normalized to allow for intensity differences in the FLIR images. Next a chip is taken from clutter II and its energy density is computed. Normalization is performed, and the chip is then classified using the discriminant function (equation 4.15).

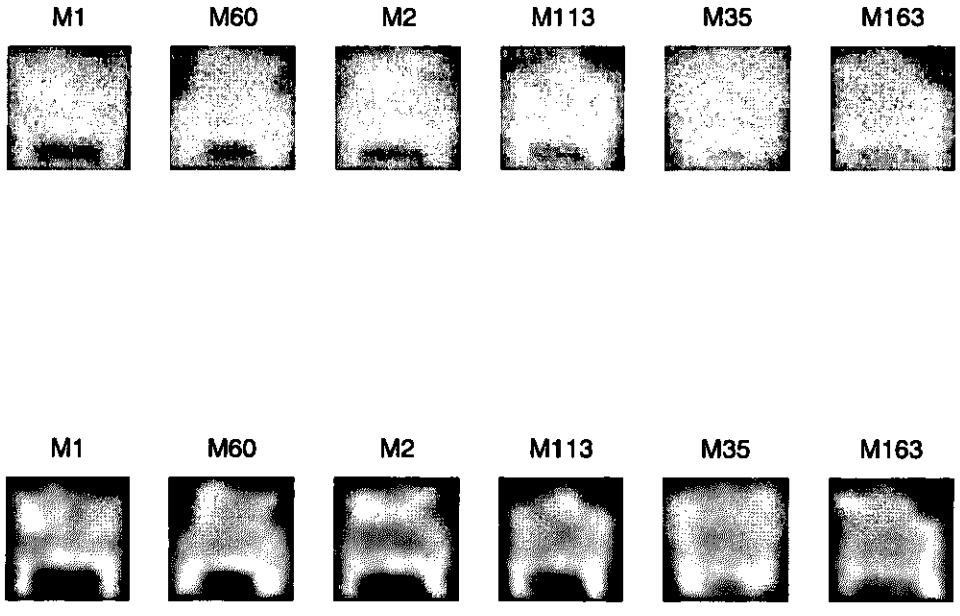

Figure 4.6: "Testing" target chips at 0° from clutter 2: (top) image space target chips; (bottom) corresponding energy density templates.

As a gauge of the performance, we repeat the experiment without taking the energy density.

4.6 Results

There are six different targets labeled M1, M2, M60, M35, M113, and M163 which are tanks, trucks and armored personnel carriers. The results of the experiment are presented in the form of a "confusion matrix," which provides information on what percentage of each class was correctly classified and the percentage that were misclassified.

Table 4.1 shows the confusion matrix that results when a classification is performed in image space, i.e., without making use of the energy density.

Table 4.1: Confusion matrix for classification without the use of the energy density.

	M1	M2	M35	M60	M113	M163
M1	97.6%	0%	2.4%	0%	0%	0%
M2	0%	97.6%	0%	2.4%	0%	0%
M35	0%	0%	100%	0%	0%	0%
M60	4.8%	0%	0%	95.2%	0%	0%
M113	0%	0%	0%	0%	100%	0%
M163	0%	0%	0%	0%	0%	100%

Table 4.2: Confusion matrix for classification using the energy density computed with the Mexican hat wavelet.

	M1	M2	M35	M60	M113	M163
M1	95.2%	0%	2.4%	0%	0%	2.4
M2	0%	100%	0%	0%	0%	0%
M35	0%	0%	100%	0%	0%	0%
M60	0%	0%	0%	100%	0%	0%
M113	0%	0%	0%	0%	100%	0%
M163	0%	0%	0%	0%	0%	100%

A 98.4% correct classification was achieved. According to the confusion matrix, about 2.5% of M1s were classified as being M35s and some M2s and M60s were also misclassified.

Table 4.2 shows the performance of the algorithm when the Mexican hat wavelet is used in the energy density calculation. In this case, although about 5% of M1s were misclassified, every other chip was identified correctly to obtain an overall classification rate of 99.2%. Table 4.3 shows that classification with the Morlet wavelet is at 95.6%, not as good as the other two cases. Only one target class was correctly classified each time.

Table 4.3: Confusion matrix for classification using the energy density computed with the Morlet wavelet.

	M1	M2	M35	M60	M113	M163
M1	97.6%	0%	0%	0%	2.4%	0%
M2	2.4%	92.8%	0%	4.8%	0%	0%
M35	0%	0%	100%	0%	0%	0%
M60	4.8%	0%	0%	95.2%	0%	0%
M113	0%	0%	0%	0%	95.2%	4.8%
M163	0%	2.4%	0%	0%	0%	97.6%

4.7 Analysis

Although the results seem to suggest certain conclusions about the relative abilities of the three methods in feature extraction, it is important to note that only 252 target chips were used, compared to over 16000 by Chan and Nasrabadi in their investigation of learning vector quantizers [4]. Hence, the results do not allow us to draw conclusively from the data. From these preliminary results, it appears that the Morlet wavelet as it was applied is not as effective at feature detection as the Mexican hat. It should be noted that we did not expect the Morlet wavelet to do as well as the Mexican hat. We were interested in detecting singularities in all directions, a task performed very well by the Mexican hat. The Morlet wavelet detection was in only the nine different directions we chose; thus it clearly approximated the function of the Mexican hat. Furthermore, the Morlet wavelet is more difficult to calibrate and it is possible that we overlooked something in the parameterization process. We had a similar problem when we chose to ignore the numerically negligible correction term and achieved classification rates below 50%. Thus it is likely that we can alter some parameters and get an improvement in performance.

An important item to note is the significance of data. Large databases are crucial for the classification part of an algorithm. Specifically, one is able to train an algorithm to capture more features extracted by the feature detector or use a different more capable classification algorithm. Chan and Nasrabadi [4] changed algorithms and achieved 5 to 10% improvement in classification. It is likely that the energy density extracted distinguishing features that could not be exploited by our classifier. The aim of the project, however, is to use our classifier in the development stage of the energy density with the view of eventually using a convolutional neural network [1] for classification.

Our next goals are to test the algorithm on a larger data set that will be acquired in the near future and expand the algorithm to recognize targets at different angles. Currently, we compare a "test" target only with templates in the "training" database at the same aspect angle. Later an "unknown" category will be included. In other words, if an input target does not match any of the templates, it should be classified as unknown.

4.8 Conclusions

We have proposed using the CWT as a means of extracting target features in an ATR classifier as an efficient and accurate technique of capturing target signatures in cluttered images. To date our results have been hindered by a lack of available data. Subsequently we can only suggest that the CWT energy density for feature detection may provide advantages over existing standard techniques. A larger data set is essential before more conclusive statements can be made.

Future efforts should include creating target-adapted wavelets that better match the type of target being analyzed. The wavelets could conceivably be shaped like the target or features of the target. A more immediate addition

would be the inclusion of some of the other energy densities such as the scale-angle energy density, whose respective features could be combined to create more detailed signatures for a target class.

References

[1] S. D. Der and R. Chellappa, "Probe-Based Automatic Target Recognition In Infrared Imagery," in *IEEE Trans. Image Processing*, Vol 6 no. 1, pp. 92-102, (1997). 1997.

[2] R. Murenzi, L. Kaplan, F. Mujica, and S. Der, "Feature Extraction in FLIR Imagery Using the Continuous Wavelet Transform (CWT)," Unpublished.

[3] L. Chan, N. Nasrabadi, and V. Mirelli, "Automatic Target Recognition Using Modularly Cascaded Vector Quantizers and Multilayer Perceptrons," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Vol. 6, pp. 3387-3390, Atlanta, (1996).

[4] L. Chan, N. Nasrabadi, and V. Mirelli, "Multi-stage Target Recognition using Modular Vector Quantizers and Multilayer Perceptrons," in *Proc. Computer Vision Pattern Recognition*, Vol. 1, pp. 114-119, San Francisco, (1996).

[5] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," in *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 17, no. 8, pp. 790-799, (1995).

[6] S. Mallat, W.L. Hwang, "Singularity Detection and Processing with Wavelets," in *IEEE Transactions on Information Theory*, Vol 38, No. 2, pp. 617-643, (1992).

[7] J-P. Antoine and R. Murenzi, "Two Dimensional Directional Wavelets and the Scale-Angle Representation," in *Signal Processing* Vol. 52, pp. 259-281, (1996).

[8] J-P. Antoine, P. Carrette, and R. Murenzi, "Image Analysis with Two-Dimensional Continuous Wavelet Transform", in *Signal Proc.*, Vol. 31, pp. 241-272, (1993).

[9] J-P. Antoine, P. Carrette, and R. Murenzi, "The Scale-Angle Representation in Image Analysis with 2D Wavelet Transform," in *Signal Processing VI: Theory and Applications (Proc. EUSIPCO 92, Brussels)*, pp. 701-704, Elsevier, Amsterdam (1992).

[10] J-P. Antoine, K. Bouyoucef, P. Vandergheyst, and R. Murenzi, "Target Detection and Target Recognition using 2 Dimensional Continuous Isotropic and Anisotropic Wavelets," in *Automatic Object Recognition V*, SPIE-1995, Vol. 2485, pp. 20-31, (1995).

[11] H. Andrews, "Introduction to Mathematical Techniques in Pattern Recognition," *Wiley Intersciences*, New York, 1972.